JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1991, 24, 293–304 NUMBER 2 (SUMMER 1991)

EVALUATION OF COMPONENTS OF RESIDENTIAL TREATMENT BY MEDICAID ICF-MR SURVEYS: A VALIDITY ASSESSMENT

DENNIS H. REID, MARSHA B. PARSONS, CAROLYN W. GREEN, AND MAUREEN M. SCHEPIS

WESTERN CAROLINA CENTER, MORGANTON, NORTH CAROLINA

We evaluated the proficiency of the federal Medicaid program's survey process for evaluating intermediate care facilities for the mentally retarded. In Study 1, an observational analysis of active treatment during leisure times in living units suggested that these surveys did not discriminate between certified and noncertified units. In Study 2, a reactivity analysis of a survey indicated that direct-care staff performed differently during the survey by increasing interactions with clients and decreasing nonwork behavior. Similarly, results of Study 3 showed increases in client access to leisure materials during a survey. In Study 4, questionnaire results indicated considerable variability among service providers' opinions on the consistency, accuracy, and objectivity with which survey teams determine agency standard compliance. Results are discussed regarding effects of the questionable proficiency of survey processes and the potential utility of behavioral assessment methodologies to improve such processes.

DESCRIPTORS: Medicaid, assessment, program evaluation, institutions

One of the most significant factors affecting residential service provision for persons with mental retardation in the United States is the Title XIX Medicaid Reimbursement Program for Intermediate Care Facilities for the Mentally Retarded (ICF-MR). Since the mid 1970s, the ICF-MR program has provided standards of care and treatment for persons with mental retardation, affecting service delivery in 49 of 50 states (Sparr & Smith, 1990) and involving about 140,000 persons with mental retardation (Holburn, 1990b). ICF-MR funds allocated to agencies based on the agencies' observed compliance with ICF-MR standards have provided a very substantial portion of operating budgets for residential facilities. Over $4.6 billion are provided annually through the ICF-MR program to participating agencies, making the ICF-MR program the largest source of funds among all federal programs for persons with developmental disabilities (Harrington & Swan, 1990).

An integral part of the ICF-MR program is the individual agency survey process ("NAPRR Submits," 1989). Although variations occur across regions of the country, the basic survey process entails at least an annual on-site review of an agency's level of compliance with the ICF-MR standards. The reported level of compliance with these standards is based on the outcome of the survey, which in turn determines the agency's continued participation in the ICF-MR program and whether or not the agency receives Medicaid funding.

In light of the importance of an ICF-MR survey outcome, it seems paramount that these surveys reach an accurate decision regarding an agency's compliance with the ICF-MR standards. In order to reach an accurate decision, a survey should be conducted in a manner that permits the accurate review of information concerning an agency's services. Some preliminary evidence suggests that ICF-MR surveys may not be very proficient in this respect. For example, Repp and Barton (1980) reported that the process appeared to have essentially no impact on the amount of habilitative programming provided to clients in one residential facility. Although habilitative programming is a primary focus of the ICF-MR standards ("NAPRR Submits," 1989), Repp and Barton found no differences in amount of programming between certified and noncertified units.

Appreciation is expressed to Jim Favell and Ed Konarski for their helpful comments on an earlier draft of the paper and to Mary Keller for her competent help in preparing the manuscript.

Correspondence and requests for reprints should be addressed to any of the authors at Western Carolina Center, Morganton, North Carolina 28655.

The purpose of this investigation was to provide a more comprehensive evaluation of the adequacy of selected components of the ICF-MR survey process than has been reported to date. In Study 1, we conducted a comparison of certified and noncertified living units in public residential facilities with regard to provision of active treatment during group leisure periods. The comparison was based on the hypothesis that because current ICF-MR standards clearly stress the provision of active treatment (Sparr & Smith, 1990), including during clients' traditional leisure times (Saunders & Spradlin, 1991), clients in certified units would spend more time in treatment-related activities and less time in nonadaptive activities (e.g., stereotypic behavior) than clients in noncertified units. In Studies 2 and 3, an analysis of the potential reactivity of an ICF-MR survey was conducted. If surveys are reactive in terms of changes in staff members' routine work performance, then it would be difficult for a survey team to determine what typically occurs in a facility and, consequently, to reach an accurate decision regarding a facility's compliance with the ICF-MR standards during routine service provision. In this regard, some data suggest that external agency surveys are quite reactive (Bible & Sneed, 1976; Quilitch, de Longchamps, Warden, & Szczepaniak, 1977). However, research to date has focused on surveys announced to an agency regarding the forthcoming date of the surveys, so that the agency knows exactly when the surveys will occur. Researchers have suggested that unannounced surveys, such as those conducted by ICF-MR survey teams, would resolve problems with reactivity (Bible & Sneed, 1976; Quilitch et al., 1977). Research has not been reported, however, on unannounced surveys, and it is unclear whether such surveys are actually nonreactive. Finally, for Study 4, we conducted a national survey of residential service providers to obtain their opinions on the accuracy with which ICF-MR survey decisions are reached.

STUDY 1

Background

In an investigation designed to demonstrate a method of improving the provision of active treatment in residential settings for persons with developmental disabilities, an initial sample of normative data from 22 living units was provided regarding levels of active treatment services (Parsons, Cash, & Reid, 1989). The sample was subsequently expanded to include a more comprehensive representation of public residential facilities in the United States, with the latter investigation involving 53 living units (Reid & Parsons, 1989). These two investigations used the same procedures to collect the normative data. Both investigations also included data collected in certified and noncertified living units. However, an analysis of the 53 living units in the latter report was not conducted regarding the differences or similarities between certified and noncertified living units. Our purpose here was to reanalyze the observational data summarized in the earlier work to compare the two types of living units. Also, additional units were observed using the same procedures employed previously to increase the sample size, particularly in regard to noncertified living units (which represented a very small minority of the living units observed in the first two studies).

Method

Participants and setting. In total, 60 living units were observed, representing 16 public residential facilities in 11 states. Fifty-three of the units had been observed in at least one of the previous two investigations just noted. An additional seven living units (three certified and four noncertified) were subsequently observed. The four noncertified units had previously been certified but, based on an ICF-MR survey during the preceding year, were judged not to be in compliance with the active treatment standard. In total, 12 of the units were not certified. These 12 units were located in two facilities, each in a different state.

In each living unit, observations of ambulatory individuals with severe or profound mental retardation were conducted. The observations occurred when groups of clients were present in the living units at times when formal off-unit day treatment programs (e.g., school, vocational services) were not in effect. Generally, these periods represented traditional leisure times (e.g., just before and after

supper) when the primary responsibility for client services fell on direct-care staff instead of specialty therapy personnel. Observations were conducted with this type of client population at these times because the provision of active treatment typically is more difficult in these situations relative to situations involving more highly skilled clients or when clients participate in formal day treatment services conducted by specialty therapy staff (Reid & Parsons, 1989). Further, based on our review of ICF-MR survey reports conducted in facilities in the southeast, midwest, and southwest regions of the country, as well as reports of others (Saunders & Spradlin, 1991), these periods present particular problems for agencies in terms of complying with the ICF-MR standards. These periods also have received more attention on recent ICF-MR surveys, due in part to changes in the ICF-MR standards that more heavily emphasize *continuous* active treatment (Saunders & Spradlin, 1991).

Behavior definitions. The primary dependent measure focused on client activity in terms of participation, or lack thereof, in active treatment pursuits. The specific behavioral definitions were drawn directly from previous work that developed and socially validated criteria for active treatment participation in schools (Green et al., 1986; Parsons, Schepis, Reid, McCarn, & Green, 1987; Reid et al., 1985) and residential settings (Parsons et al., 1989; Reid & Parsons, 1989). Given the focus of this paper, attention was directed to only two categories of resident activity: active treatment and nonadaptive behavior.

Active treatment was defined as client involvement in an activity with an apparent purpose, such as working independently on an habilitative task, manipulating leisure materials, interacting with a staff member, or receiving assistance from a staff member (e.g., being physically guided through an habilitative task). If a client was manipulating materials, then the materials had to be used in the manner for which they were intended. If the materials were not being used in the manner for which they were intended (e.g., stacking toy trucks on top of each other), a client was considered to be engaged in an activity, but not as a part of active treatment. Active treatment excluded resident self-care routines such as getting dressed or being dressed by a staff person, watching television and aggressive or distruptive behavior.

Nonadaptive behavior was defined by exclusion as lack of client involvement in any of the behaviors just noted. As alluded to earlier, nonadaptive activity was characterized by stereotypic behavior, nonpurposeful wandering, and sitting or lying on the floor with no apparent activity. Aggressive or disruptive behavior was not included here because it was considered to be a mutually exclusive category in previous work to determine its specific occurrence frequency and the corresponding need for immediate intervention and because it previously was observed very infrequently, during no more than 3% of all observation intervals (Parsons et al., 1989).

Observation system. The observation system was identical to that used previously (for elaboration see Parsons et al., 1989; Reid & Parsons, 1989). Briefly, at a given time a systematic count was made of the number of clients involved in active treatment pursuits and the number involved in nonadaptive activity. An observer entered a living unit (e.g., dayroom, activity room) and, using a left-to-right view of the area, quickly identified the clients present. Beginning with the first client listed, the observer then watched the client only long enough to determine what the client was doing when first observed (no more than 5 s). Five additional seconds were allowed to record the observed activity, and then the next client on the list was observed. All remaining clients were observed sequentially in this manner until each client had been observed for at least two samples, or until a minimum of 10 samples was obtained across all clients. For most living units (73%), observations were conducted for 1 or 2 days, whereas for the remainder the observations took 3 to 7 days. In total, there were 127 separate observations (averaging 2.1 observations per living unit).

Interrater reliability was evaluated by two persons observing simultaneously and independently. Reliability checks were conducted on 49% of all observations involving certified and noncertified living units. Agreement percentages were calculated using an interval-by-interval system (Bailey & Bos-

tow, 1979) to determine occurrence, nonoccurrence, and overall reliability by dividing the number of agreements by the number of agreements plus disagreements and multiplying by 100. Overall reliability averaged 95% for nonadaptive behavior (range, 67% to 100%) and 97% for active treatment (range, 77% to 100%). Occurrence reliability averaged 92% (range, 50% to 100%) and 77% (range, 0% to 100%), respectively, and nonoccurrence averaged 86% (range, 40% to 100%) and 95% (range, 70% to 100%), respectively. The lower figures typically were attributable to low frequencies of opportunities to observe, resulting in a small number of disagreements deflating the average.

Results and Discussion

There was no apparent difference in the mean percentage of clients observed to be involved in active treatment between certified (M = 17%, SD = 13) and noncertified (M = 20%, SD = 10) units. Similarly, there was no difference in nonadaptive behavior (M = 72% for both types of units; SD = 16 for certified units and 12 for noncertified units). Median levels of active treatment and nonadaptive behaviors were also very similar for certified and noncertified units. The median level of active treatment involvement was 21% for noncertified units and 16% for certified units, and was 72% for both types of units for nonadaptive behavior. Aggressive or disruptive behavior was very infrequent, averaging less than 2% of all observation intervals.

Results of Study 1 suggest the ICF-MR certification and review process may not reliably discriminate between ICF-MR certified and noncertified units based on relative amounts of active treatment during leisure times in living units. The lack of such discrimination is of particular concern when considering the noted ICF-MR emphasis on continuous active treatment, including during leisure times in living units, for an agency to obtain and maintain ICF-MR certification (Holburn, 1990a). Of course, one explanation for the lack of observed differences across certified and noncertified units is that the definition of active treatment and the observation system used here do not represent a valid or sufficiently comprehensive set of criteria for evaluating active treatment, at least from an ICF-MR perspective. It also should be noted that the sample of certified and noncertified living units was not randomly selected. However, both types of units served the same general type of population—ambulatory persons with severe or profound mental retardation. Additionally, the noncertified living units were in the same two states as the certified units observed, and both types of living units were part of the same state-level agency operations (although in separate residential facilities). The observed staff-to-client ratios likewise were similar, although somewhat greater in the noncertified units (overall average of 2.5 staff to 8.9 clients per unit) than the certified units (1.9 to 9.8).

The important qualifications just noted notwithstanding, the criteria for active treatment used in this investigation have received support from a broad range of professionals (see the social validity section of Parsons et al., 1989), and the observation system has been demonstrated to be reliably sensitive to comprehensive changes in staff and supervisory performance that increase active treatment (Parsons et al., 1989). In addition, the sample of living units observed (N = 60) was considerably larger than samples in typical observational studies of living units (Blindert, 1975; Harmatz, 1973; Repp & Barton, 1980). To our knowledge, no other approach to defining, measuring, and evaluating active treatment in residential settings has received the degree of validation and application research as the system and sample used in this investigation.

STUDY 2

Background

As noted earlier, the purpose of Study 2 was to evaluate whether the ICF-MR survey process represents a reactive procedure in terms of different staff behavior (from an active treatment perspective during leisure periods in living units) during an ongoing ICF-MR survey relative to routine staff behavior. An opportunity arose to evaluate survey

reactivity during a separate investigation (Green, Reid, Perkins, & Gardner, in press). Systematic measures were being collected on staff performance to establish a baseline prior to implementing a staff management program to improve staff performance. During baseline data collection, an ICF-MR survey was conducted. Baseline measures continued after the survey was completed, resulting in a retrospective quasi-experimental design (Agras, Jacob, & Lebedeck, 1980) with which to evaluate the effects of the survey on staff behavior.

Method

Setting and participants. The setting was one living unit in an ICF-MR certified residential facility serving persons with developmental disabilities. The staff participants were 4 direct-care personnel assigned to the afternoon shift of the living unit. Biographical and educational characteristics of the 3 women and 1 man were similar to those representative of direct-care personnel in other state residential facilities (Ford, 1983). These staff members were selected because they constituted all of the full-time direct-care personnel on the afternoon shift of that living unit. The client population for the living unit consisted of 20 adults who had profound multiple handicaps (Landesman-Dwyer & Sackett, 1978). The clients were profoundly mentally retarded (Grossman, 1983), nonambulatory, and had serious medical complications such as frequent seizure activity as well as multiple physical disabilities (e.g., quadriplegia). All clients essentially were totally dependent on staff for fulfillment of basic care and treatment needs.

Behavior definitions. Several types of staff behavior were observed. However, of concern here are two behavior categories: *interaction activities* and *nonwork behavior.* Interaction activities were defined as a staff member engaging in a social or leisure interaction with a client (e.g., talking to a client about his or her clothes, singing a song, playing a musical instrument for a client, helping a client activate a mechanical apparatus such as a tape recorder). Nonwork behavior was defined as a staff member not engaging in any behavior related to living environment maintenance or client welfare, such as watching television, reading a magazine, talking to other staff persons about nonwork activities, and sitting in a chair with no other apparent activity. These two categories were targeted for two reasons. First, interaction activities represented a desired staff behavior based on management directives at the time. Second, from a management perspective, nonwork behavior was the least desired staff behavior.

Observation system. From 4:00 to 4:30 p.m., an observer monitored the behavior of staff members at 5-min intervals. Observations involved a time sampling system in which an observer entered the living unit, located each staff person present (1 to 4 staff members were present at any given time), noted the behavior of each staff person when first observed, and recorded the number of staff members engaged in each type of behavior. Reliability checks were conducted during the second and seventh observation sessions (of a total of nine observations) by two observers observing simultaneously and independently. There were no disagreements between observers regarding the occurrence and nonoccurrence of each behavior category.

Experimental conditions—baseline. During baseline, staff engaged in their usual routines. They had been instructed to encourage client leisure activity and to interact frequently with clients during this period. They had also received repeated training sessions regarding how to interact with clients and how to train clients to enhance participation in leisure activities.

Experimental conditions—ICF-MR survey. The ICF-MR survey was conducted by the state ICF-MR office in a manner that had occurred at least annually at the facility for a number of years. The survey, which involved an on-site review of client records and periodic observations in program areas (e.g., schools) and living units, began on Monday and ended on Friday. The time of the survey was not known to the staff prior to the beginning of the survey on Monday, at which time an announcement of the survey team's presence was made. The survey team visited the unit in which experimental observations were made but not during these observations.

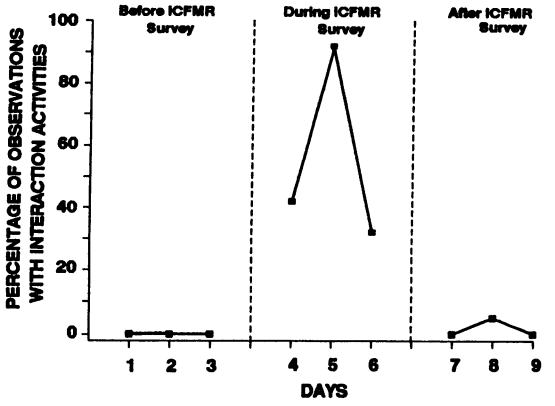

Figure 1. Percentage of observations in which staff members were involved in interaction activities for each day during each experimental condition.

Experimental Design

The experimental design consisted of a retrospective, quasi-experimental A-B-A reversal (Agras et al., 1980). Specifically, baseline (A) was represented by the ongoing work activities of staff during 3 routine weekdays, followed by the work activities during the external survey conducted by the state ICF-MR office (B), and then followed in turn by 3 routine workdays without an external survey (A).

Results and Discussion

The effects of the ICF-MR survey on staff behavior are reflected in Figure 1. Observations conducted before the survey indicated that there were no interaction activities implemented by staff during the target period. In contrast, during the survey, staff-conducted interaction activities occurred during an average of 55% of the observations of staff behavior. Subsequently, following the survey interaction activities decreased to an average of 2%. Nonwork behavior changed in the opposite direction, averaging 19% during the first baseline, 7% during the survey, and 35% during the second baseline.

Given the nature of the observation system (i.e., recording the number of staff involved in each type of behavior per time sample), a detailed analysis of individual staff behavior is not possible. However, the group averages presented in Figure 1 are representative, at least generally, of changes in individual staff performance. That is, it appears that the increases in interactions during the week of the ICF-MR survey were the result of changes in performance among all, or at least 3 of the 4, staff members across all observations relative to both baseline conditions. During the initial baseline, no observation of interactions occurred. During the second baseline, only 1 staff person was observed to interact (hence, the average level of interactions for the other three staff persons had to be 0%). In contrast, in the observations made during the survey, 1 to 4 staff persons were observed to be interacting; therefore, the average level of interactions for all staff persons was much higher than 0%.

Due to the retrospective, quasi-experimental nature of the investigation, the results should be interpreted with some caution. Although reliability observations were being conducted as part of the initial investigation (Green et al., in press) and occurred during the first and second baselines in this study, because of the retrospective nature of this study it could not be ensured that the reliability checks included all experimental conditions. However, additional support for reliability stems from the more frequent checks involving the same observers, behavior definitions, and setting as part of the initial study (Green et al.). Further, use of single-subject designs in a quasi-experimental fashion has precedent in applied behavior analysis (Agras et al., 1980; McSweeny, 1978) when preplanned applications of more formal experimental designs cannot be performed because of ethical or other reasons. In this case, preplanned evaluations of ICF-MR surveys would be very difficult because ICF-MR suvey teams are not allowed to announce forthcoming survey dates to an agency.

Results of Study 2 suggest that the ICF-MR survey process was quite reactive. Staff members interacted much more frequently with clients while the survey was taking place and spent less time in nonwork behavior relative to their more routine work pattern. These results suggest that the ICF-MR surveyors did not observe a typical leisure time period for the clients of this particular living unit.

Consequently, it would be difficult for the ICF-MR survey team to form an accurate judgment regarding service provision based on the survey and to make valid recommendations regarding compliance with the ICF-MR certification standards. However, these results and corresponding conclusions should be qualified in light of the small sample of staff persons observed. An opportunity arose to evaluate potential reactivity of the ICF-MR survey with a larger sample during a survey that occurred 2 years after the survey described in Study 2.

STUDY 3

Background

As in Study 2, an evaluation of the potential reactivity of an ICF-MR survey became possible when a survey occurred shortly after baseline measures had begun in two different living units of the same residential facility described in Study 2. However, instead of observing staff interactions and nonwork behavior, the focus of Study 3 was the availability of leisure materials in the living units for client use. The importance of leisure material availability for promoting appropriate leisure behavior and discouraging maladaptive behavior for persons with severe handicaps has been discussed elsewhere (Risley & Favell, 1979). The dependent measures were part of a pilot project to train new direct-care staff members in methods of providing a therapeutic environment for clients. One component of training involved systematic observations in the two living units in which the staff worked. When the ICF-MR survey occurred, observers were recording the availability of leisure materials in the living room area of the unit.

Method

Setting and participants. The two living units each served 32 clients who were ambulatory and had profound or severe mental retardation. The observations occurred during the afternoon shift. The direct-care staff assigned to the units were similar to those described in Study 2 in terms of biographical and educational characteristics. There were 10 afternoon staff members assigned to both Unit 1 and Unit 2, with 5 to 7 typically present on a given day in each unit.

Dependent variable and observation process. The dependent variable of interest was the percentage of clients who had a leisure or habilitative material within arm's reach. A leisure material was defined as any item not unusual for a nonhandicapped person to use during leisure time (Parsons et al., 1989); typical examples are magazines, radios, coloring paper and crayons, puzzles, and various toys. An habilitative material was defined as any item usable in a functional skill training program (Parsons et al., 1989), including make-up kits and adaptive switch mechanisms.

The observation process consisted of a brief time sampling system. Upon entering the living room area, the observer recorded demographic information (place, time, number of clients present, etc.) and then noted if a leisure or habilitative material was within arm's reach of each client when first observed. Subsequently, the total number of clients with material access was divided by the total number of clients present and multiplied by 100 to derive a percentage figure. Observations were conducted during traditional leisure times (i.e., generally between 3:30 and 6:00 p.m. or 7:00 and 8:00 p.m.) when the majority of the clients were present in the living units. As in Study 2, the ICF-MR survey team visited both units, although not at the exact times of observations. Reliability observations occurred as described previously, during the fourth observation in Unit 1 and during the second and third observations in Unit 2. Across all reliability checks, observers never disagreed on the number of clients who had materials within arm's reach.

Experimental conditions and design. The experimental conditions were the same as in Study 2 in terms of work days before, during, and after an ICF-MR survey. The general living unit routines and prior training histories of the staff were very similar to those in Study 2. During leisure times, it was the responsibility of the staff to distribute leisure materials to the clients from various storage

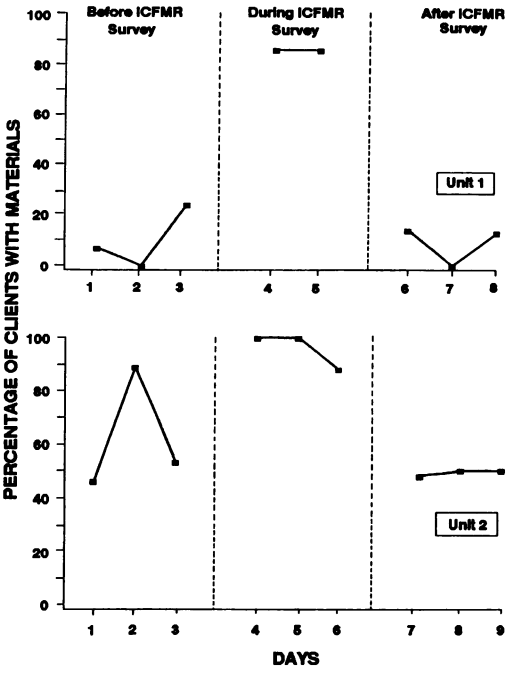

Figure 2. Percentage of clients with immediate access to leisure materials for each day during each experimental condition for Living Unit 1 (top panel) and Living Unit 2 (bottom panel).

compartments. The experimental design (a quasi-experimental A-B-A) was the same as in Study 2.

Results and Discussion

As indicated in Figure 2, the percentage of clients with leisure or habilitative materials in immediate access increased in both units during the survey. In Unit 1 (top panel of Figure 2), the percentage of clients increased from a baseline mean of 14% to 86% during the survey and then decreased to 9% after the survey. Similarly, respective averages for Unit 2 (bottom panel) were 63%, 96% and 49%, although there was more variability during the initial baseline in Unit 2.

Results of Study 3 support results of Study 2 in suggesting that staff performed differently during an ICF-MR survey relative to their usual work performance. In this case, staff appeared to provide more leisure and habilitative materials to clients during the survey. Again, however, because of the quasi-experimental nature of the investigation, the same qualifications offered in Study 2 are appropriate here. Also, the measure of staff performance was more indirect in that it focused on the presumed outcome of staff performance (i.e., materials made available to clients). Nevertheless, clients themselves did not have access to the materials without staff assistance, because the materials typically were placed in storage cabinets. Without changes in staff performance it is very unlikely that the changes in material access that corresponded to the presence and absence of an ICF-MR survey could have occurred. Such changes, involving a different client-related measure as well as more and different staff than in Study 2, lend external validation to results of Study 2.

STUDY 4

Studies 1 through 3 were based on systematic observations of service provision in relation to ICF-MR survey processes and decisions. In contrast, Study 4 involved a more subjective, opinion-based evaluation format to provide a comprehensive, multifaceted evaluation of ICF-MR surveys.

Method

Participants and settings. Questionnaires were sent to 243 directors of residential services in state-operated programs serving persons with mental retardation located in every state in the United States. The facilities were selected from the *Directory of Public Residential Facilities for the Mentally Retarded* (1985).

Opinion questionnaire. The questionnaire consisted of six questions. The first question asked respondents to indicate whether or not their program was at least partially ICF-MR certified. Only individuals whose facilities participated in the ICF-MR program were asked to respond to the remaining five questions (Table 1), which assessed opinions regarding the ICF-MR surveys with which the residential directors were familiar. The directors were asked to respond to each question on a 5- or 6-point Likert-type scale.

Procedure. The questionnaires were mailed to the directors, along with a brief cover letter indicating an interest in obtaining service providers'

Table 1
Distribution of Service Providers' Responses to Opinion Questionnaire

How important do you believe it is for your facility to be in compliance with ICF-MR standards?				
75% extremely important	18% very important	3% minimally important	1% very unimportant	2% extremely unimportant
From survey to survey, how consistently do survey teams evaluate your facility in regard to the teams' interpretation of your facility's compliance to the ICF-MR standards?				
2% extremely consistent	29% very consistent	49% somewhat consistent	19% very inconsistent	2% extremely inconsistent
How well or poorly do you believe the existing method of conducting an ICF-MR survey results in an accurate differentiation between high quality treatment services versus low quality treatment services?				
1% extremely well	27% very well	52% somewhat well	16% very poorly	5% extremely poorly
How do ICF-MR survey decisions appear to be made in regard to your facility's compliance with ICF-MR standards?				
1% extremely objective	24% very objective	51% somewhat objective	22% very subjective	3% extremely subjective
Overall, how helpful or detrimental is the ICF-MR survey process to your agency in regard to improving resident services?				
7% extremely helpful	33% very helpful	49% somewhat helpful	10% somewhat detrimental	2% very detrimental	0% extremely detrimental

opinions of the ICF-MR survey process. Approximately 8 weeks after the initial mailing, the questionnaire was mailed again to individuals who had not yet responded, in an effort to prompt additional responses (Winett, Stewart, & Majors, 1978). A stamped, self-addressed envelope was included with both mailings.

Results and Discussion

A total of 195 questionnaires were returned, representing an 80% return rate. Results of the questionnaire indicated considerable variability in the opinions of service providers on the ICF-MR survey process (Table 1). Regarding the importance of complying with the ICF-MR survey standards, almost all (93%) of the respondents indicated that the process was at least very important. Most respondents (89%) also indicated that the process was at least somewhat helpful for improving client services. However, in contrast to these rather favorable opinions of the survey process, only 31% of the respondents reported that survey teams' interpretation of their program's compliance with the ICF-MR standards was very (or extremely) consistent across different surveys, with the most common response (49%) reporting that the surveys were only somewhat consistent. Further, 21% reported ICF-MR surveys to be very or extremely inconsistent.

Approximately one fifth (21%) of the respondents also reported that the surveys do a poor job of differentiating between high- and low-quality service provision, whereas a slightly greater percentage (28%) reported the opposite. With respect to the degree of objectivity or subjectivity with which survey decisions appear to be made regarding a program's compliance with standards, 51% reported that decisions were only somewhat objective, and as many respondents (25%) reported the decisions to be made very or extremely subjectively as did the respondents who reported the decisions to be made very or extremely objectively. Overall, these results indicate that although service providers who are experienced with the ICF-MR process believe ICF-MR compliance is important and the

survey process is helpful, there is considerable disagreement regarding the degree of consistency, accuracy, and objectivity of ICF-MR teams' interpretations of an agency's compliance with the ICF-MR standards.

GENERAL DISCUSSION

The results of this investigation suggest that there are some serious problems with specific components of the existing survey process of the ICF-MR program. One problem was reflected in Study 1, in which the survey process was not discriminative across certified versus noncertified living units when based on a direct observational evaluation of active treatment designed to be objective, systematic, and socially valid. These results coincide with observational data reported on a much smaller scale by Repp and Barton (1980), although the focus of the two observational systems differed in terms of specific ICF-MR variables addressed.

A second problem with the ICF-MR survey process was noted in Studies 2 and 3, in which staff behavior appeared to be very reactive to the presence of a survey, at least in regard to the staff variables monitored. This reactivity suggests that conditions during a survey are not very representative of routine service provision. These results indicate that the reactivity problems noted earlier with external surveys announced to agencies prior to the actual survey, such as sanitation surveys by state health departments (Quilitch et al., 1977) and program evaluations by the Joint Commission on Accreditation of Hospitals (Bible & Sneed, 1976), are not necessarily resolved with unannounced visits.

In Study 4, a third problem was revealed by the highly divergent opinions of service providers on the proficiency of the ICF-MR survey process. Many service providers reported considerable inconsistency and subjectivity with the ICF-MR survey process or outcome, although at least as many service providers did not express problems in this regard. The results of the nationwide survey, which had a very high response rate (80%), corroborate other reports of service providers' concerns regarding problems with the survey process ("NAPRR Submits," 1989).

The focus of this investigation addressed only one component of the ICF-MR survey process, namely that of active treatment provision. Also, the focus of Studies 1, 2, and 3 was on only one component of active treatment (i.e., living units in contrast to formal day-treatment settings). Hence, suggested problems with the survey process should not be generalized to other components. Nonetheless, active treatment in living units represents a substantial part of the current ICF-MR standards as well as a part with which many facilities recently have experienced problems with compliance (Saunders & Spradlin, 1991). The results obtained here cast doubt on the proficiency with which certification processes related to this component of active treatment services are conducted and, subsequently, at least part of the basis on which certification decisions are made. Such results have serious practical implications. In particular, as noted earlier, a very large amount of federal money is expended through the ICF-MR program, and it is not clear that very proficient processes are used to determine which agencies receive or relinquish those funds. Relatedly, perceptions of subjectivity and inconsistency that exist among many service providers regarding ICF-MR survey processes and decisions can result in considerable disgruntlement among service providers ("NAPRR Submits," 1989).

Based on the pervasive impact of the ICF-MR program and the results of the four experiments, continued research is warranted to examine further the proficiency, or lack thereof, regarding the methodology and outcome of the entire ICF-MR survey process. Research would also be useful on methods of improving the proficiency of components of the process. Such research could address, for example, increasing the objectivity and consistency of methods of evaluating active treatment provision through the use of behavioral assessment methodologies (Cone & Hawkins, 1977), perhaps along the lines of the observation system in Study 1. This type of assessment would consist of behavioral definitions of client involvement in active treatment, a systematic time sampling observation process, and inde-

pendent reliability checks among surveyors. Behavioral definitions and systematic observations also may be used to help control for the reactivity of the external survey process. That is, if the definitions and observational process were clearly described in replicable terms, the information could be disseminated to service providers. The service providers could then use the same assessment procedures and periodically send their assessment data to the ICF-MR personnel for review. External surveyors could then compare their own data with those of the service providers to check for, among other things, staff reactivity to the survey (i.e., by determining whether data collected during the survey by ICF-MR staff present a more favorable picture of an agency's services relative to the data provided routinely by the service agency). If agencies frequently conduct data-based observations using the same processes used by external review agencies, the increased frequency of the observations itself may serve to reduce the reactivity of the external surveys (i.e., the external surveys may then become less of a novel or unusual event). Such a process might also provide a more representative evaluation base for ICF-MR teams because it could result in more evaluative data than an ICF-MR survey conducted only once or twice a year.

The proposed solutions to some of the problems with ICF-MR surveys as just noted are overly simplified, and more detailed discussion is warranted. Nonetheless, technically sound behavioral assessment methodologies do exist to help governmental regulatory and funding bodies, such as ICF-MR, be more systematic and objective in their assessment procedures. In essence, if such regulatory bodies review human service agencies with the expectation that the agencies' service provision is in line with state-of-the-art standards, then it seems appropriate to conduct the regulatory review processes in accordance with state-of-the-art assessment methodology.

REFERENCES

Agras, W. S., Jacob, R. G., & Lebedeck, M. (1980). The California drought: A quasi-experimental analysis of social policy. *Journal of Applied Behavior Analysis,* **13,** 561–570.

Bailey, J. S., & Bostow, D. E. (1979). *Research methods in applied behavior analysis.* Tallahassee, FL: Copy Grafix.

Bible, G. H., & Sneed, T. J. (1976). Some effects of an accreditation survey on program completion at a state institution. *Mental Retardation,* **14,** 14–15.

Blindert, H. D. (1975). Interactions between residents and staff: A qualitative investigation of an institutional setting for retarded children. *Mental Retardation,* **13,** 38–40.

Cone, J. D., & Hawkins, R. P. (Eds.). (1977). *Behavioral assessment: New directions in clinical psychology.* New York: Brunner/Mazel.

Directory of public residential facilities for the mentally retarded. (1985). Fairfax, VA: National Association of Superintendents of Public Residential Facilities for the Mentally Retarded.

Ford, J. E. (1983). Application of a personalized system of instruction to a large, personnel training program. *Journal of Organizational Behavior Management,* **5**(3/4), 57–65.

Green, C. W., Reid, D. H., McCarn, J. E., Schepis, M. M., Phillips, J. F., & Parsons, M. B. (1986). Naturalistic observations of classrooms serving severely handicapped persons: Establishing evaluative norms. *Applied Research in Mental Retardation,* **7,** 37–50.

Green, C. W., Reid, D. H., Perkins, L. I., & Gardner, S. M. (in press). Increasing habilitative services for persons with profound handicaps: An application of structural analysis to staff management. *Journal of Applied Behavior Analysis.*

Grossman, H. J. (Ed.). (1983). *Classification in mental retardation.* Washington, DC: American Association on Mental Deficiency.

Harmatz, M. G. (1973). Observational study of ward staff behavior. *Exceptional Children,* **39,** 554–558.

Harrington, C., & Swan, J. H. (1990). State Medicaid ICF-MR utilization and expenditures in the 1980–1984 period. *Mental Retardation,* **28,** 15–27.

Holburn, C. S. (1990a). Rules: The new institutions. *Mental Retardation,* **28,** 89–94.

Holburn, C. S. (1990b). Symposium overview: Our residential rules—have we gone too far? *Mental Retardation,* **28,** 65–66.

Landesman-Dwyer, S., & Sackett, G. P. (1978). Behavioral changes in nonambulatory, profoundly mentally retarded individuals. In C. E. Meyers (Ed.), *Quality of life in severely and profoundly mentally retarded people: Research foundations for improvement* (pp. 55–144). Washington, DC: American Association on Mental Deficiency.

McSweeny, A. J. (1978). Effects of response cost on the behavior of a million persons: Charging for directory assistance in Cincinnati. *Journal of Applied Behavior Analysis,* **11,** 47–51.

NAPRR submits first compilation of responses to HCFA "spotters" program. (1989, September). *Living in New Kinds of Situations,* pp. 9, 11.

Parsons, M. B., Cash, V. B., & Reid, D. H. (1989). Im-

proving residential treatment services: Implementation and norm-referenced evaluation of a comprehensive management system. *Journal of Applied Behavior Analysis,* **22,** 143–156.

Parsons, M. B., Schepis, M. M., Reid, D. H., McCarn, J. E., & Green, C. W. (1987). Expanding the impact of behavioral staff management: A large-scale, long-term application in schools serving severely handicapped students. *Journal of Applied Behavior Analysis,* **20,** 139–150.

Quilitch, H. R., de Longchamps, G. D., Warden, R. A., & Szczepaniak, C. J. (1977). The effects of announced health inspections upon employee cleaning performance. *Journal of Organizational Behavior Management,* **1**(1), 79–88.

Reid, D. H., & Parsons, M. B. (1989). An objective and systematic method for reviewing ongoing active treatment in congregate living units. *Superintendent's Digest,* **8,** 103–106.

Reid, D. H., Parsons, M. B., McCarn, J. E., Green, C. W., Phillips, J. F., & Schepis, M. M. (1985). Providing a more appropriate education for severely handicapped persons: Increasing and validating functional classroom tasks. *Journal of Applied Behavior Analysis,* **18,** 289–301.

Repp, A. C., & Barton, L. E. (1980). Naturalistic observations of institutionalized retarded persons: A comparison of licensure decisions and behavioral observations. *Journal of Applied Behavior Analysis,* **13,** 333–341.

Risley, T. R., & Favell, J. (1979). Constructing a living environment in an institution. In L. A. Hamerlynck (Ed.), *Behavioral systems for the developmentally disabled: II. Institutional, clinic, and community environments* (pp. 3–24). New York: Brunner/Mazel.

Saunders, R. R., & Spradlin, J. E. (1991). A supported routines approach to active treatment for enhancing independence, competence and self-worth. *Behavioral Residential Treatment,* **6,** 11–37.

Sparr, M. P., & Smith, W. (1990). Regulating professional services in ICF's/MR: Remembering the past and looking to the future. *Mental Retardation,* **28,** 95–99.

Winett, R. A., Stewart, G., & Majors, J. S. (1978). Prompting techniques to increase the return rate of mailed questionnaires. *Journal of Applied Behavior Analysis,* **11,** 437.

Received June 27, 1990
Initial editorial decision October 18, 1990
Revision received December 19, 1990
Final acceptance February 19, 1991
Action Editor, John M. Parrish